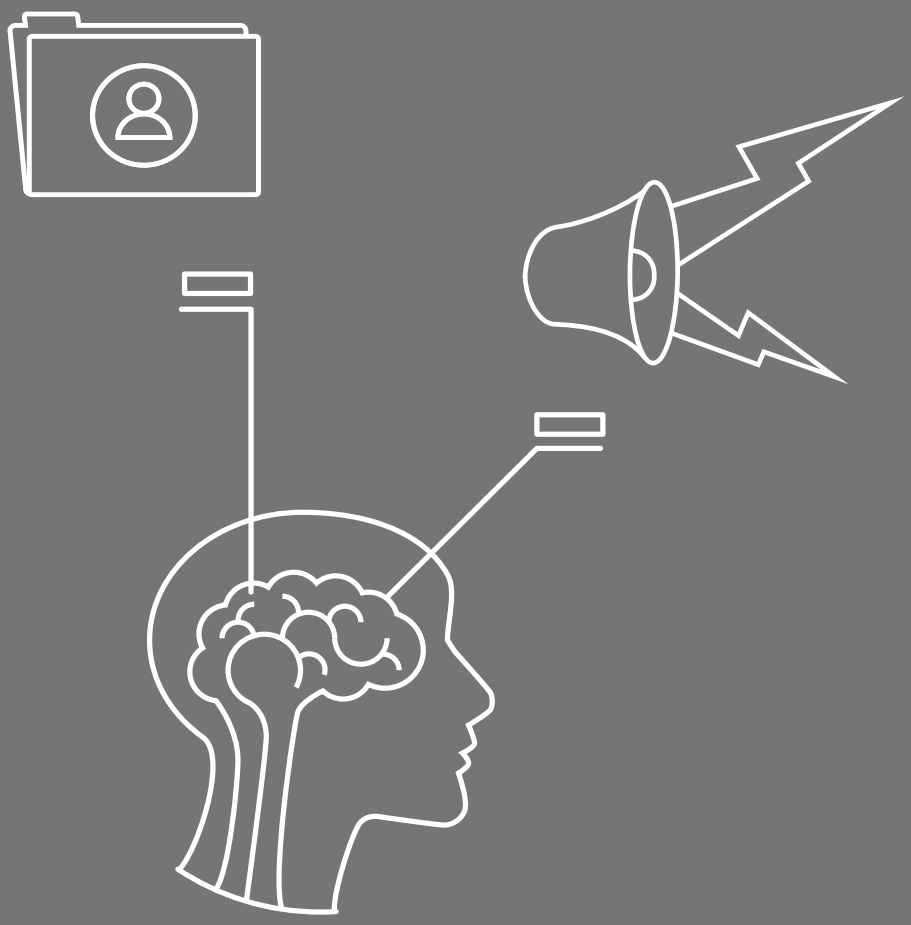


Memes That Kill: The Future Of Information Warfare

CBINSIGHTS

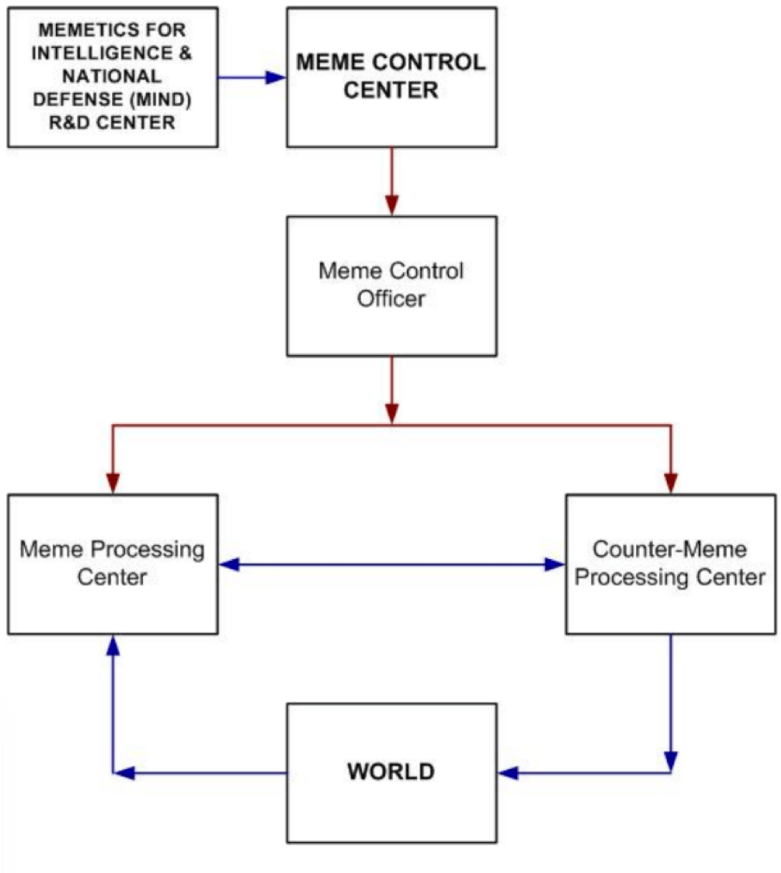
Memes and social networks have become weaponized, while many governments seem ill-equipped to understand the new reality of information warfare. How will we fight state-sponsored disinformation and propaganda in the future?

In 2011, a university professor with a background in robotics presented an idea that seemed radical at the time.

After conducting research backed by DARPA — the same defense agency that helped spawn the internet — Dr. Robert Finkelstein proposed the creation of a brand new arm of the US military, a "Meme Control Center."

In internet-speak the word "meme" often refers to an amusing picture that goes viral on social media. More broadly, however, a meme is any idea that spreads, whether that idea is true or false.

It is this broader definition of meme that Finklestein had in mind when he proposed the Meme Control Center and his idea of "memetic warfare."



From "Tutorial: Military Memetics," by Dr. Robert Finkelstein, presented at Social Media for Defense Summit, 2011

Basically, Dr. Finklestein's Meme Control Center would pump the internet full of "memes" that would benefit the national security of the United States.

Finkelstein saw a future in which guns and bombs are replaced by rumor, digital fakery, and social engineering.

Fast forward seven years, and Dr. Finklestein's ideas don't seem radical at all. Instead, they seem farsighted.

Memetics and the Tipping Point

- **Rumor: most contagious of all social messages**
- **Kernel of truth in rumor (or urban legend) is distorted in 3 directions:**
  - **(1) Level story (i.e., leave out details essential for understanding true meaning of incident)**
  - **(2) Sharpen story (i.e., make remaining details more specific)**
  - **(3) Assimilate story (i.e., change story so it makes sense to those spreading the rumor)**
  - **Example: rumors of 9/11**

From "Tutorial: Military Memetics," by Dr. Robert Finkelstein, presented at Social Media for Defense Summit, 2011

The 2016 US presidential election was shaped by a volatile mix of fake news, foreign meddling, doctored images, massive email leaks, and even a cartoon meme (Pepe the Frog). Not to mention a conservative news site called Infowars.

It no longer seems silly to say that the future of warfare isn't on the battlefield, but on our screens and in our minds.

Military and intelligence agencies around the world are already waging secret information wars in cyberspace. Their memes are already profoundly influencing public perceptions of truth, power, and legitimacy.

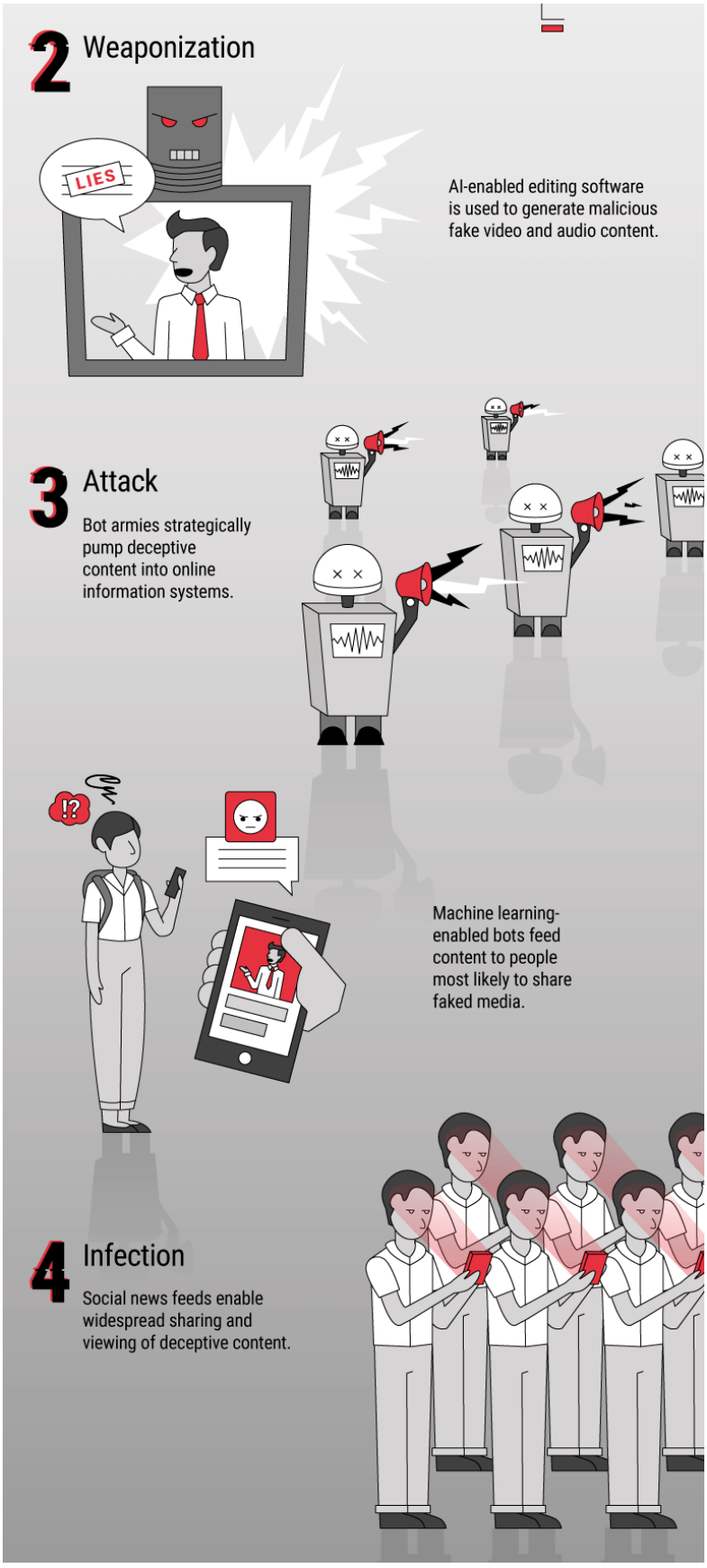
And this threat is only intensifying as artificial intelligence tools become more widely available.

Consider:

- Political-bot armies or fake user "sock puppets" are targeting social news feeds to computationally spread propaganda.
- Online, the line between truth and falsehood is looking fragile as AI researchers develop technologies that can make undetectable fake audio and video.
- Within a year, it will be extremely easy to create high-quality digital deceptions whose authenticity cannot be easily verified.

Below, we detail the technologies, tactics, and implications of the next generation of war.

FUTURE OF INFORMATION WARFARE
Wars of the future will use computational propaganda and advanced digital deceptions to distort the enemy's perception of reality and manipulate public opinion. The information warfare attack chain:
1 Reconnaissance
An adversary scrapes and steals targets' metadata.
The adversary uses metadata to create a psychographic profile to identify targets' vulnerabilities.


2 Weaponization
LIES
AI-enabled editing software is used to generate malicious fake video and audio content.
3 Attack
Bot armies strategically pump deceptive content into online information systems.
!?
Machine learning-enabled bots feed content to people most likely to share faked media.
4 Infection
Social news feeds enable widespread sharing and viewing of deceptive content.

Information attacks — like the one depicted above — can be summed up in one centuries-old word: Provokatsiya, which is Russian for "act of provocation." The act is **said** to have been practiced by spies in Russia, dating back to the late Tsarist era. Provokatsiya describes staging cloak and dagger deceptions to discredit, dismay, and confuse an opponent.

The terrorizing drums, banners, and gongs of Sun Tzu's warfare, aided by information technology ... may now have evolved to the point where 'control' can be imposed with little physical violence.

US Colonel Richard Szafranski

"A THEORY OF INFORMATION WARFARE: PREPARING FOR 2020", WRITTEN IN 1995

In addition to international interference, politicians have also been known to stage domestic digital influence campaigns. President Trump's campaign has come under increasing scrutiny for **reportedly** contracting UK-based firm Cambridge Analytica to mine Facebook data and influence voter behavior in the run-up to the 2016 election.

However, we focus on cases of a foreign adversary attacking another country (as opposed to domestic influence campaigns), and on state-sponsored acts of information warfare (as opposed to acts perpetrated by unaffiliated actors).

Table of contents

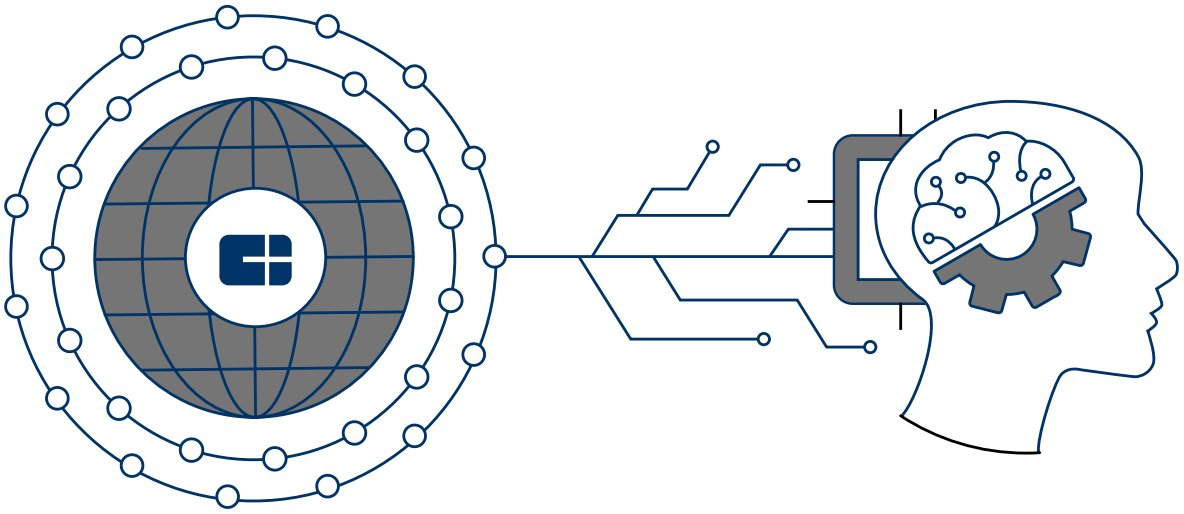
1 The rise of digital information warfare

9 Key elements of the future of digital information warfare

- Diplomacy & reputational manipulation
- Automated laser phishing
- Computational propaganda

18 Emerging solutions in the fight against digital deception

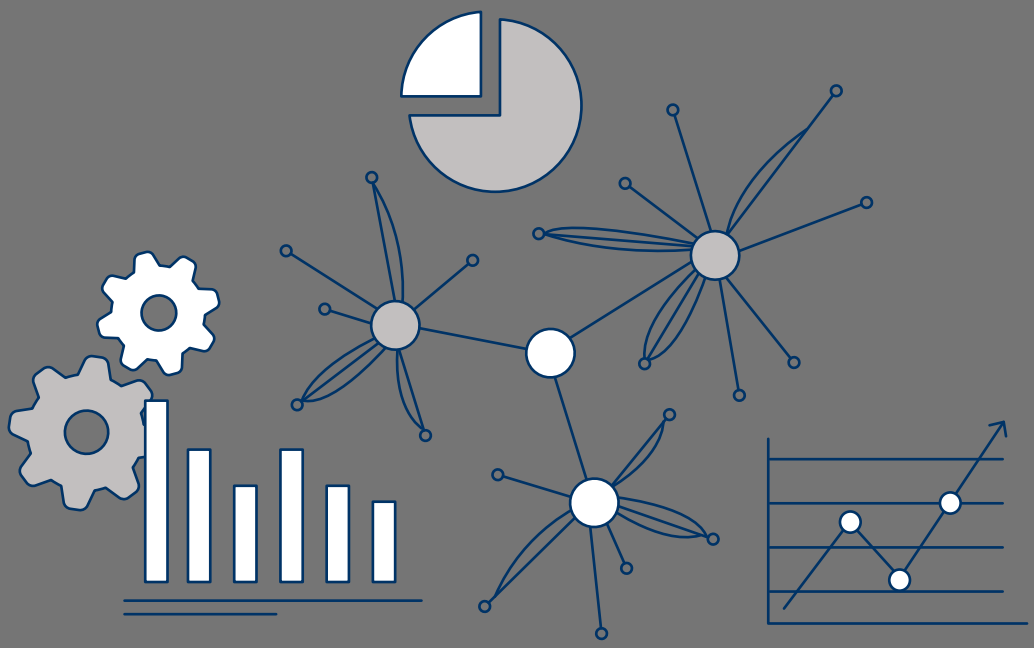
- Uncovering hidden metadata for authentication
- Blockchain for tracing digital content back to the source
- Spotting AI-generated people
- Detecting image and video manipulation at scale
- Combating computational propaganda
- Government regulation & national security
- Final thoughts



At CB Insights, we believe the most complex strategic business questions are best answered with facts.

We are a machine intelligence company that synthesizes, analyzes and visualizes millions of documents to give our clients fast, fact-based insights.

From Cisco to Citi to Castrol to IBM and hundreds of others, we give companies the power to make better decisions, take control of their own future—and capitalize on change.



WHERE IS ALL THIS DATA FROM?

The CB Insights platform has the underlying data included in this report

CLICK HERE TO SIGN UP FOR FREE

"We use CB Insights to find emerging trends and interesting companies that might signal a shift in technology or require us to reallocate resources."

Beti Cung,
CORPORATE STRATEGY, MICROSOFT

Microsoft

TRUSTED BY THE WORLD'S LEADING COMPANIES

citi IBM GE BARRICK Castrol
SEQUOIA DASSAULT SYSTEMES CISCO salesforce

1

The rise of digital information warfare: how did we get here?

Generally, information wars involve two types of attacks: acquiring **sensitive data and strategically leaking it**, and/or waging **deceptive public influence campaigns**.

Both types of attacks have made waves in recent years. In one of the most notorious examples, Russian agents staged informtion attacks intended to influence the outcome of the 2016 US presidential election. Russian cyber troops **reportedly** hacked and leaked sensitive email communications from the Democratic National Committee and conducted an online propaganda campaign to influence American voters.

Facebook **agrees** with the FBI's **indictment** that a Russian government contracted unit called the Internet Research Agency (IRA) was responsible for exposing up to 150M Americans (or two-thirds of the electorate) to foreign propaganda via the social media platform. The indictment does not say whether Russia's meddling had an effect on the election's outcome. But the electoral and media system's vulnerability is a worry for everyone, regardless of partisan politics.

Of course, not all information leaks are clear acts of war. In some cases, leaks serve as a stepping stone toward accountability and transparency as is now considered the case with the so-called Pentagon Papers that revealed the extent of the US secret war in Southeast Asia.

Essentially, leaks are a grey area. Each leak must be examined on a case-by-case basis before it is declared an act of war.

Targeted disinformation campaigns are not a grey area: they are malicious and corrosive. These attacks (including disinformation, propaganda, and digital deception) are the focus of this research.

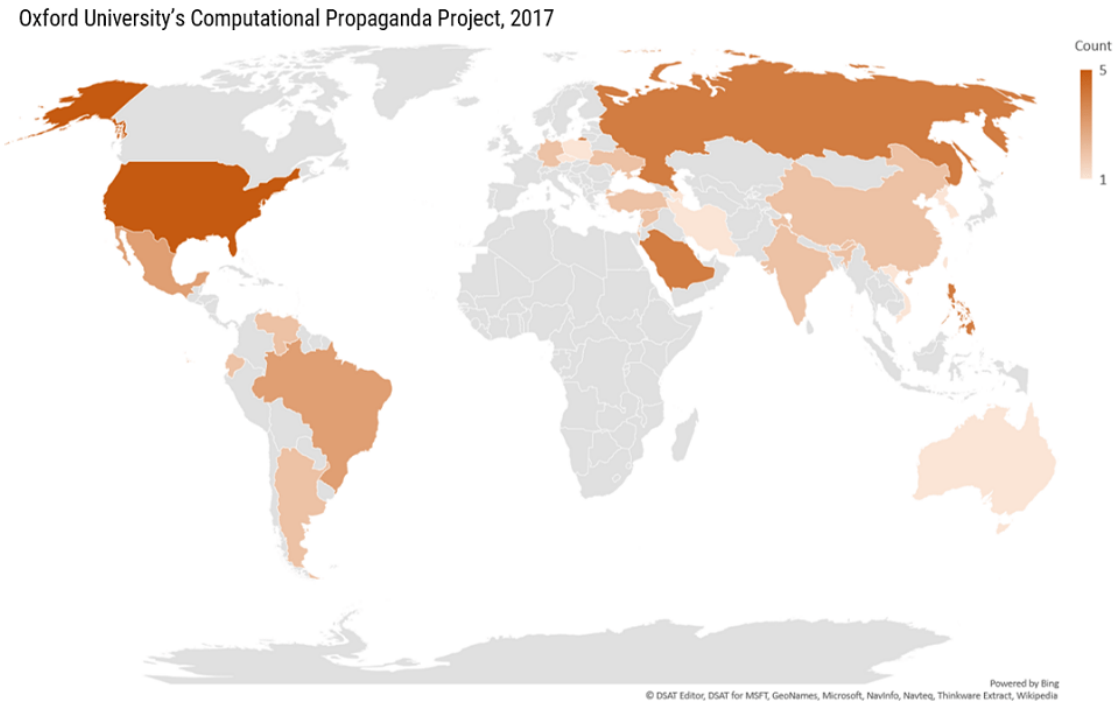
In recent years, **information attacks have materialized quickly**. Four years ago the World Economic Forum named the "spread of misinformation online" the 10th most significant trend to watch in 2014. Today, events like Russia's election meddling confirm the systematic state-sponsored deployment of digital information attacks by a foreign adversary.

In other words, in just two years (2014 — 2016) a bad actor's ability to manipulate information on the internet went from barely being a top ten concern among thought leaders to likely having a direct effect on the American democratic process.

Russia is not the only country responsible for distorting public opinion on the internet. *An Oxford University* **study** *found instances of social media manipulation campaigns by organizations in at least 28 countries since 2010*. The study also highlighted that "authoritarian regimes are not the only or even the best at organized social media manipulation".

Typically, **cross-border information wars are waged by state-sponsored cyber-troops**, of which the world has many and the US has the most.

Density of state-sponsored cyber-attack units by country


Oxford University's Computational Propaganda Project, 2017
Count
5
1
Powered by Bing
© DSAT Editor, DSAT for MSFT, GeoNames, Microsoft, Navinfo, Navteq, Thinkware Extract, Wikipedia

Source: Oxford University

The world is already facing the uncomfortable reality that people are increasingly confusing fact and fiction. However, the technologies behind the spread of disinformation and deception online are still in their infancy, and the problem of authenticating information is only starting to take shape.

Put simply, this is only the beginning.

There is no Geneva Convention or UN treaty detailing how a nation should define digital information attacks or proportionally retaliate. As new technologies spread, understanding the tactics and circumstances that define the future of information warfare is now more critical than ever.

2

Key elements of the future of digital information warfare

One common theme in digital information wars to come will be the intentional spreading of fear, uncertainty, and doubt also known as FUD online. Negative or false information will be hyper-targeted at specific internet users that are likely to spread FUD.

Three key tactics, buoyed by supporting technologies, will play key roles in the future of war, as delineated in part by Aviv Ovadya, chief technology officer for the University of Michigan's Center for Social Media Responsibility:

- **Diplomacy & reputational manipulation:** the use of advanced digital deception technologies to incite unfounded diplomatic or military reactions in an adversary; or falsely impersonate and de-legitimize an adversary's leaders and influencers.
- **Automated laser phishing:** the hyper-targeted use of malicious AI to mimic trustworthy entities that compel targets to act in ways they otherwise would not, including the release of secrets.
- **Computational propoganda:** the exploitation of social media, human psychology, rumor, gossip, and algorithms to manipulate public opinion.

Through most of history the primary purpose of military operations has been achieved through physical activity … nowadays almost all acts of physical violence come with an [online] component, exploiting social networks to manipulate opinion and perception.

General Sir Nicholas Houghton

FORMER CHIEF OF THE DEFENCE STAFF
OF THE BRITISH ARMED FORCES

Diplomacy & reputational manipulation: faking video and audio

Diplomacy manipulation is the act of creating a false belief that an event has occurred in order to influence geo-political decisions.

To that end, researchers at the University of Washington (UW) have already successfully used AI to create a realistic video of President Obama "saying" things he never actually said.

According to the university's paper on the experiment, grafting audio clips onto a realistic, lip synched video can "change what [Obama] appears to be saying in a target video to match the input audio track."

It's easy to imagine how such an altered video — if good enough to look authentic — could quickly wreak havoc, either in the US or abroad.

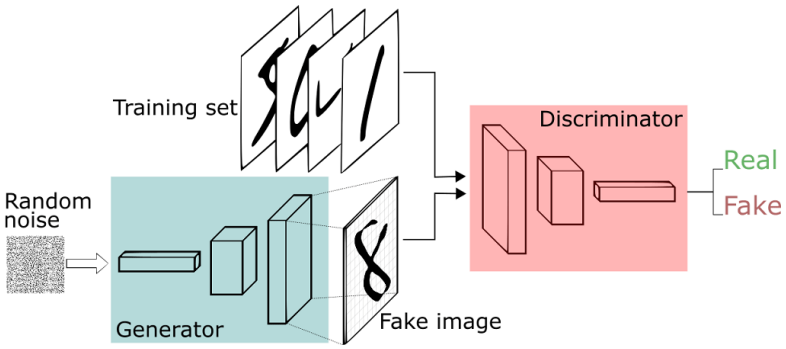
Advances in AI are ushering in a new era of fake video and audio that will have profound effects on the future of diplomacy. While the tech is largely still under development in universities, that won't be the case for long.

AI to create fake digital content

GANs (generative adverserial networks) are a type of AI used to carry out **unsupervised machine learning**. In a GAN, opposed neural networks work together to fabricate increasingly realistic audio, image, and video content.

Essentially, one neural network in the GAN acts as a foil that pushes the other network to generate more high-fidelity results. The network is judged and corrects its output until the end result is a truly realistic video or picture of an event that never actually happened.

Neural networking also makes it easier to fake audio. A neural network can convert the elements of an audio source into statistical properties, and those properties can be rearranged to make original fake audio clips.



Sketch of a General Adversarial Network for creating fake images, credit DL4J

High-caliber diplomatic manipulations will likely combine AI, audio, and video deception into one attack. Right now, research teams around the world are working on seemingly benign tools that could provide that opportunity if we are not careful.

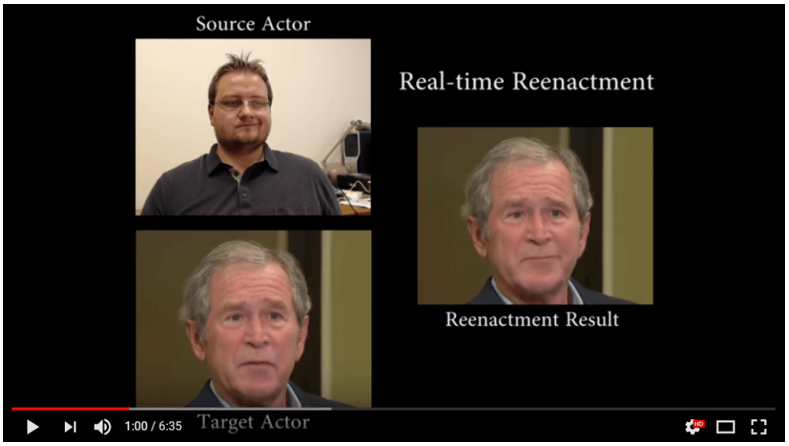
Deceptive video & audio editing

Stanford University researchers published early results indicating that it is possible to alter a person's pre-recorded face in real-time to mimic another person's expressions.

Essentially, an actor wanting to impersonate a target can create a **digital human puppet** by making a face into a webcam. A digital rendition of the target's face will mimic the actor's face in real-time.

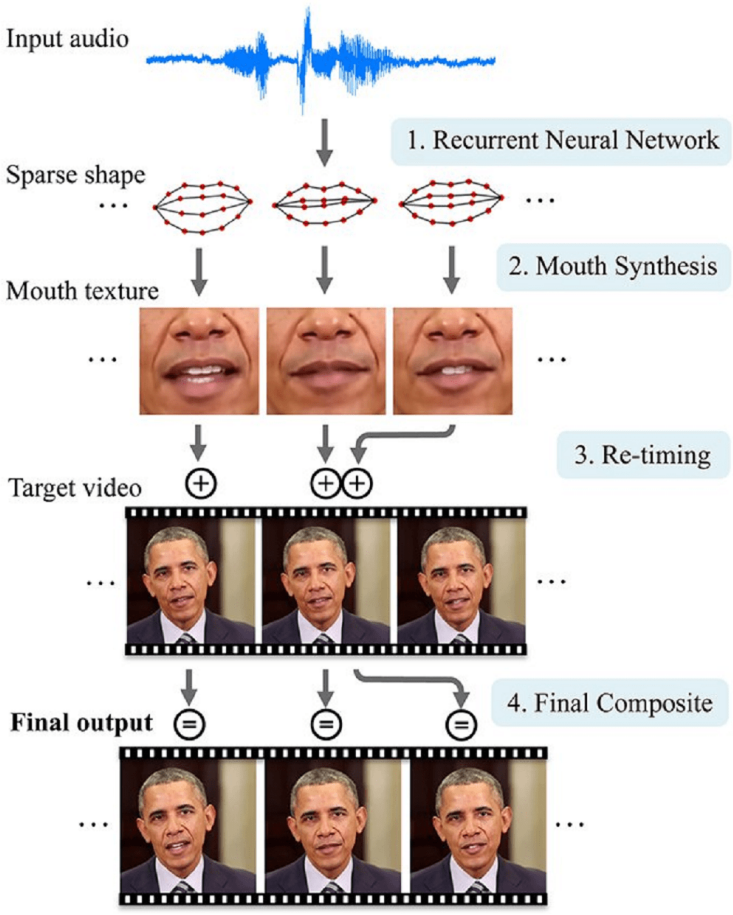
For now, this technique still requires hours of pre-recorded video footage of the target in order to look realistic. Unfortunately, powerful public figures are uniquely vulnerable, since there is ample historic video footage showing them speaking and movingin real life.

The use of a low-tech webcam along with this high-tech software suggests that the technique is accessible to video hobbyists and sophisticated propaganda artists alike. This levels the playing field, widening the gamut of actors capable of fake video deceptions.



Source: Face2Face, Stanford University

In the case of UW's Obama video experiments, which was funded by Samsung, Google, Facebook, Intel, and the university's Animation Research Labs, researchers used a neural network to first convert the sounds from an audio file into basic mouth shapes. Then the system grafted and blended those mouth shapes onto an existing target video and adjusted timing to create a new realistic, lip-synced video.



Source: University of Washington's sketch of the process that created the fake Obama video

Future iterations of the lip-synch tech being developed at UW are focused on using less data to generate the fake clips — going from 10 or more hours of video training data down to just one. If researchers are successful in doing so, the tech will be usable to create fake videos of people with less historic footage to train the algorithms.

Notably, **fabricating audio is becoming easier and more consumerized.** For example, Canadian startup Lyrebird is developing technology that can record one minute of audio from someone's voice to generate longer fabricated audio clips in the same voice. Similarly, in 2016 Adobe unveiled a prototype called Project VoCo (also dubbed "Photoshop for voice"). The project aims to let users edit human speech the same way Photoshop can be used to edit digital pictures.

These tools will undoubtedly impact the future of diplomatic decision making: imagine a rash of fake videos muddying the waters during sensitive peace negotiations in a conflict-ridden part of the world.

Meanwhile, the widespread adoption of AI-enabled video and audio meddling means there will also be a corresponding rise in reputational attacks against high-value targets.

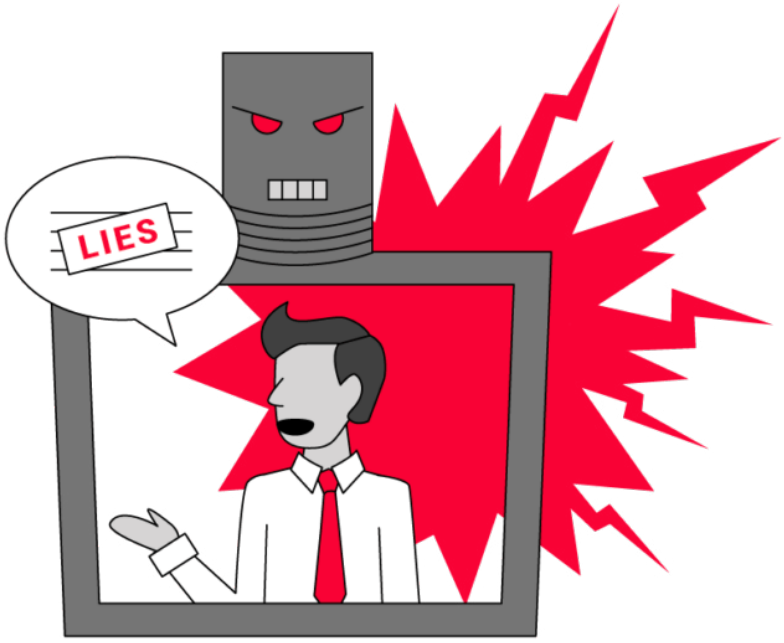
Reputational manipulation involves the use of digital video and audio deceptions to attack a person's reputation.

People have already begun to make pornographic videos, known as **deepfakes** (a portmanteau of "deep learning" and "fake"), using software that superimposes celebrities' faces onto adult film stars. The term deepfake first appeared on Reddit when an anonymous user known as "deepfakesapp" released the first version of the technology in December 2017.

Another Redditor later released an improved version, called **FakeApp**. FakeApp uses a deep learning program called TensorFlow, developed by Google, to allow users to create realistic videos where faces have been swapped.

The app and its underlying technology are gaining traction.

Deepfake Society, a website that curates deepfake videos made using FakeApp, has had over 1 million views since launching in February. Deepfake Society bans pornography (it has sister sites that do not). Similar sites have sprung up as free-resources for acquiring the tools and skills necessary to perform rudimentary deepfake operations such as grafting former Vice President Joe Biden's face onto a video of President Trump.



Source: DeepFakesClub

Reputational attacks can defame a person's character, render them untrustworthy, and even create a case for arrest. In war, this tactic will be used to discredit leaders and inflame societal tensions.

Reputational warfare technology is still fairly new, used primarily by early adopters in relatively remote corners of the internet.

The technology is progressing alongside advancements in AI that can take into account convincing details such as eye movements, wrinkles, dimples, and more.

I'm much more worried about what could come next — could bad actors target kids with fake videos from people they trust?

Senator Mark Warner
(D – VA)

Both the source code and the entire FakeApp project with pre-trained models can be found online and open-source on GitHub. The software teaches itself to perform image-recognition tasks through trial and error. The more computer processing power, the faster it works.

A looming danger with reputational attacks like deepfakes is not just that unassuming people will believe in hoaxes, but also that people will be able to **dismiss video and audio evidence of true crimes.**

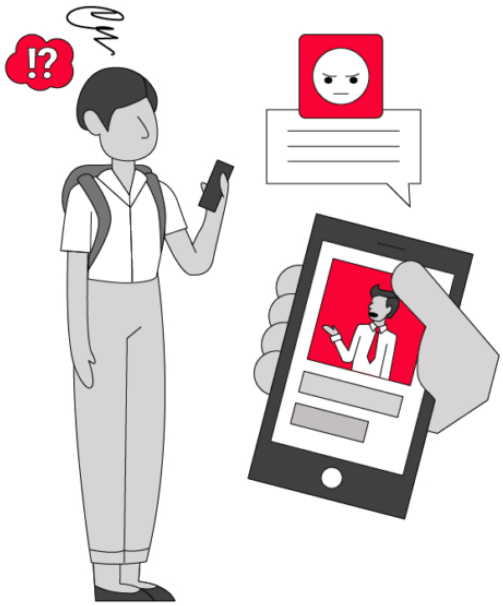
Automated laser phishing: Malicious AI impersonating and manipulating people

The amount of data that is available these days makes individuals vulnerable to a multitude of personal attacks. Common cyber attacks known as **phishing** will be the primary means of waging personalized attacks — and such attacks are getting increasingly sophisticated and difficult to stop.

Automated laser phishing attacks **use AI to create realistic impersonations of people**. The intent of these attacks is to coerce others into taking certain actions and/or divulging secret information.

Spear-phishing perpetrated through email is the most common form of targeted cyber attack — and the incorporation of AI means attackers will get better at selecting, impersonating, and fooling their victims. AI-enabled phishing makes victims much more likely to trust attackers, while automation accelerates the scale at which attacks can occur.

The sophistication and scale of these attacks means entire populations can be fooled into following the lead of a malicious AI. A torrent of disinformation could come from an AI impersonating key decision makers, causing widespread confusion.



Alarmism can be good — you should be alarmist about this stuff… We are so screwed it's beyond what most of us can imagine. We were utterly screwed a year and a half ago and we're even more screwed now. And depending how far you look into the future it just gets worse.

Aviv Ovadya

CHIEF TECHNOLOGIST FOR THE UNIVERSITY OF MICHIGAN'S CENTER FOR SOCIAL MEDIA RESPONSIBILITY

The combination of reputation attacks and automated laser phishing will essentially make it so we can't trust what we see and hear from other people online.

Over time — and with enough exposure to these kinds of digital deceptions — this can result in **reality apathy**.

Reality apathy is characterized by a conscious lack of attention to news and a loss of informedness in decision-making. In the US, an increasingly uninformed electorate could hurt the premise of our democracy, while in authoritarian states, monarchs could further entrench their control over uninformed and apathetic citizens.

Computational propaganda: digitizing the manipulation of public opinion

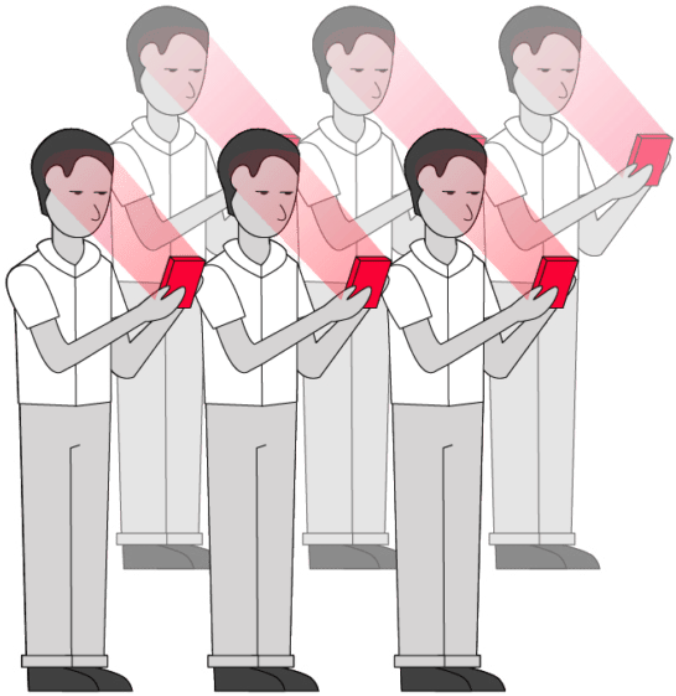
Computational propaganda is the use of algorithms, automation, social media, and human-curated content to wage widespread public influence campaigns.

Social media is vital to the flow of computational propaganda. The algorithms that curate our social news feeds are susceptible to manipulation. Social news feeds are governed by incentives that prioritize extreme views and shareable content over quality and truth.

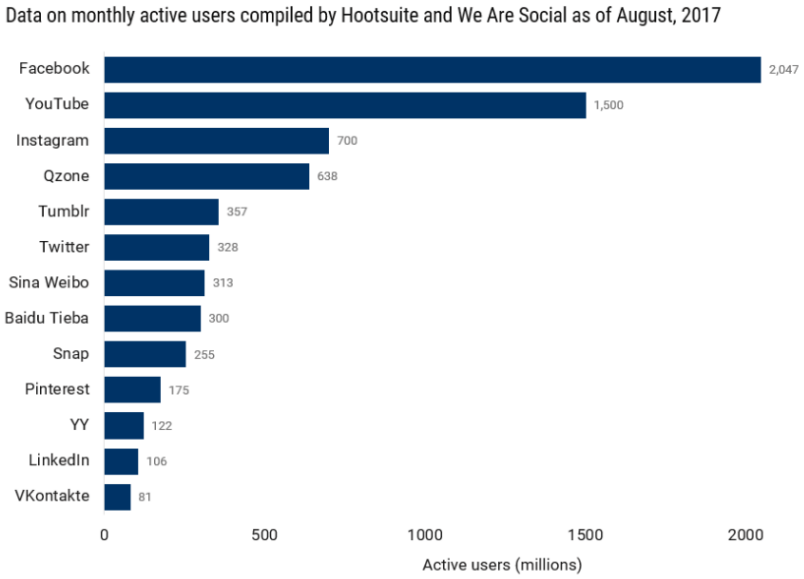
This creates a scenario in which users ingest and promote narratives that they believe to be true, regardless of their validity.

Many of the most influential tech companies either own or back the social networking platforms that host the world's computationally distributed propaganda.

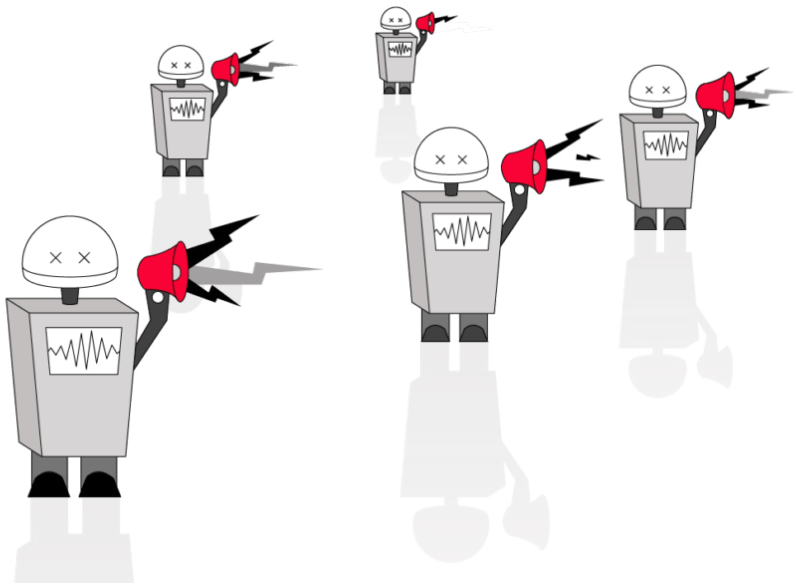
Facebook is the most widely used social-network, followed by Google's YouTube, and then Facebook-owned Instagram. Tencent in China owns the platform Qzone, in fourth.



Key social platforms ranked by number of users



The problem is exacerbated by the fact that **40%** of the global population uses social media. In the US, more than two-thirds of Americans (67%) get at least some news on social media, according to a 2017 Pew Research Center **study**.



Democracy depends on an informed electorate, and when we can't even agree on the basics of what's real, it becomes increasingly impossible to have the hard conversations necessary to move the country forward… The cumulative effect of this is a systemic erosion of trust, including trust between people and their leaders.

Renee DiResta

POLICY LEAD AT DATA FOR DEMOCRACY

And social media use isn't constrained to developed nations: a Pew Research Center **study** of 21 developing and emerging nations found that people in advanced economies use social media daily for news purposes at similar rates to those in emerging or developing economies (median of 36% and 33%, respectively).

However, America's trust in the mainstream media has steadily declined since the early 2000s, with less than half of the country **indicating** that they trust major media institutions in 2017.

Notably, **bots are integral to the spread of computational propaganda**. Bots are software programs designed to mimic humans. Security experts believe that bots generate just over half (**~52%**) of all online traffic.

On social networking platforms, bots make it so that one person, or a small group of people, can falsely give the impression that large-scale social and political movements exist where they in fact do not.

Computational propaganda bots are used in large-scale mining of a target population's metadata. They then manipulate that metadata to identify the right digital channels to flood with propaganda, with a goal of pushing out perfectly timed targeted information that is aimed at specific users.

AI and machine learning technologies enable computational propaganda bots to tailor their campaigns in real-time and spread with virus-like scale. Essentially, these bots identify and exploit people who are computationally pre-determined to be the most vulnerable to digital psychological manipulation.

Political propagandists, such as the infamous firm Cambridge Analytica, **exploit traits in people** that signal their level of susceptibility to different psychological manipulations.

Examples of such traits are detailed in leaked emails from the (now **bankrupt**) firm. Traits include allegiance to a political party, stances on hot-button issues such as gun-control, and even if a person is neurotic, suspicious of others, or believes in astrological signs.

Dr Alex Kogan [redacted] 9 May 2014 at 02:14
To: Christopher Wylie [redacted]

Hey Chris and [redacted],

Great chatting with you both yesterday. Here is a list of traits that can be predicted now--good starting shopping list. More specific items can also be predicted within the bigger personality questionnaires, but this is a start to get you thinking about what you may want:

- openness
- conscientiousness
- extraversion
- agreeableness
- neuroticism
- life satisfaction
- iq
- gender
- age
- political views = conservative?
- political views = liberal?
- political views = uninvolved?
- political views = libertarian?
- religion (categorical)
- job (categorical)
- university subject concentration (categorical)
- self-disclosure (do you tell people about yourself or not?)
- fair-mindedness (fair or suspicious in dealings with others?)
- self-monitoring (do you change your personality depending on who you're with)
- sensational interests (has 5 factors, "militarism" [guns and shooting, martial arts, crossbows, knives], "violent occultism" [drugs, black magic, paganism], "intellectual activities" [singing and making music, foreign travel, the environment], "credulousness" [the paranormal, flying saucers], "wholesome interests" [camping, gardening, hill-walking], see: https://www.academia.edu/157864/The_first_sensational_interests_paper - it's used in forensic psychology to understand criminality)
- belief in star signs (5 point scale)

Shopping list of predictable traits compiled by Cambridge Analytica, source: NYT

The future of computational propaganda

In the future, **memetic warfare** and computational propaganda will go hand-in-hand.

Memetic warfare might be seen as the digital-native version of traditional psychological warfare. As we stated above, **memes** are seemingly benign forms of digital media that spread, often as mimicry or for humorous purposes, on social networks.

More broadly, memes have proven able to derail or ignite political campaigns, polarize people, fuel social movements, and even incite violence.

The concept of memetic warfare has been around in military circles since at least the mid-2000s. In fact, as we stated above, DARPA commissioned research on "military memetics" as part of its dive into "neurocognitive warfare" as early as 2006.

In the near future, it is likely that we will see computationally spread memetic warfare campaigns that are sophisticated enough to reprogram individuals' views on a scale that manipulates the behavior of entire societies.

We could also see the rise of **dark-web computational propaganda marketplaces**, where vulnerable peoples' information will be grouped and sold off to the highest bidder.

These will be illicit, anaonymous markets for **buying and selling profiles of psychologically malleable people**. Adversarial governments and foreign agents can use these markets to gain over-the-counter access to vulnerable parties.

In all this, it's important to note that acquiring the metadata needed to run nationwide psychographic modeling is expensive and time-consuming. Creating the initial templates for deceptive content and training the computational models is at least in part the work of human teams who require pay.

Therefore, computational propaganda campaigns are likely going to be waged by the rich and powerful against the rich and powerful: governments, political parties, corporations, and special interests groups.

3

Emerging solutions in the fight against digital deception

There is no time to wait for a solution. Countries including Egypt, Brazil, and Mexico all have general elections in 2018, and in the US, 2018 midterm elections are around the corner. These political races and many others will be increasingly manipulated by computational propaganda and advanced digital deceptions.

We must develop new technologies and techniques to combat information warfare. To begin, we need a scalable way to spot high-quality fake videos.

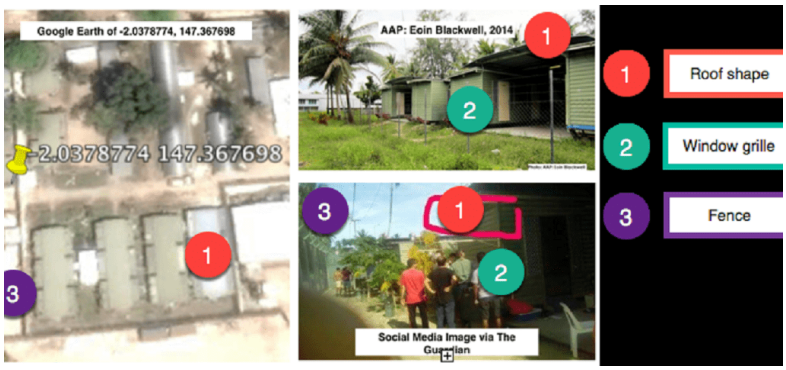
Putting a stop to computational propaganda is an even more complex problem. Nevertheless, researchers and professionals in universities, governments, startups, and the nonprofit sector are laying the groundwork for what could someday become effective forensic defenses in the fight against digital deception.

Uncovering hidden metadata to authenticate images and videos

Amnesty International, the world's largest grassroots-funded human-rights organization, is on the front lines in the fight to authenticate user-submitted video evidence of human rights abuses.

Amnesty's Citizen Evidence Lab specializes in **uncovering the context behind images and videos**. The lab is building expertise and technology to authenticate when, where, and even how a video was captured.

For example, the lab uses Google Earth and the search engine Wolfram Alpha to cross-reference surroundings and weather conditions in videos to see if the video was captured under the conditions it claims.



Citizen Evidence Lab triangulates details in user-submitted video of a shooting in Papua New Guinea to authenticate the video's origins.
Source: Amnesty International

The Citizen Evidence lab also has a tool called the YouTube Data Viewer, which **extracts hidden metadata** from videos hosted on YouTube. Most of the work centers on identifying old or forged videos that users try to pass off as current human rights abuses.

Blockchain for tracing digital content back to the source

Cryptographic techniques that underpin the technology behind blockchain can also help ensure that digital content comes from a trusted, accountable source.

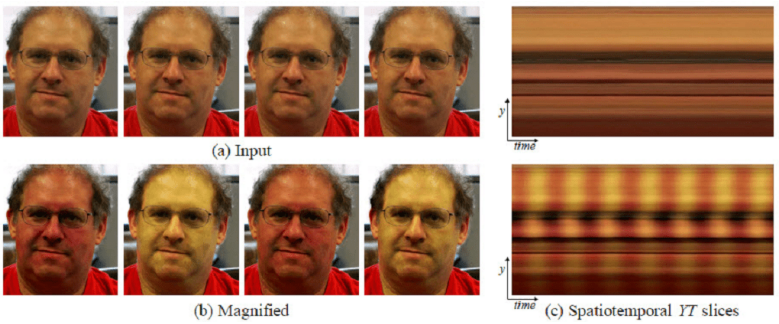
Essentially, media could be stamped with a unique cryptographic identifier, which — when cross-referenced with records on a blockchain — can prove beyond a doubt where the media originated. Media without an identifier would be considered less trustworthy.

This technique would be especially helpful for spotting images and videos that are used out of context in attempts to deceive. However, digital forensics groups and media-cryptographers will still have to grapple with AI generating fake videos from scratch.

Spotting AI-generated people

Researchers at MIT have **demonstrated** Eulerian Video Magnification technology that can help spot AI-generated people in videos.

This video magnification technology can identify real vs AI-generated people by detecting minute details such as a person's heart rate by looking at subtle changes in skin color due to blood flow. By detecting the absence of facial-blood-flow, the technology can detect computer-fabricated subjects: AI is not yet good enough to create that level of realism in a fake video.



Source: MIT

The research is a first step in the fight to distinguish footage of real people from digital ghosts.

Cross referencing video-metadata, documenting legitimate content on the blockchain, and using advanced video magnification technology are steps in the right direction. However, these tools and techniques are not scalable enough to eradicate the looming threat of open-source AI enabled digital deceptions. We need a scalable solution for the day when almost anyone can make a high-quality fake.

Detecting image and video manipulation at scale

The Defense Advanced Research Projects Agency (DARPA) has launched at least two calls for research to build a scalable digital media authentication system

The Media Forensics (MediFor) project is an attempt to build a platform for **algorithmically detecting manipulations in images and videos**. MediFor could one day lead to the creation of a crowdsourcing platform where viewers can collectively investigate videos' authenticity.

DARPA's MEMEX project could help build a massive online search engine capable of **cross-referencing image data from the entire internet**, including the deep web. One MEMEX-funded project from Columbia University demonstrates the ability to find similar images of human-trafficking victims amongst terabytes of structured and unstructured data. That work could help uncover aspects of AI-generated images and videos that originated from other sources.

Combatting computational propaganda

We can defend ourselves from being algorithmically manipulated on social media. The key is to spot digital propaganda in time to take action.

Promising work is being done to develop algorithms that **collect and categorize instances of digital propaganda to identify bots and accounts** that are responsible for its spread. However, at present a scalable technological means of countering computational propaganda is largely theoretical, and at best in the early stages of development.

Several organizations have published ideas on how to combat computational propaganda, such as leveraging **good bots that disrupt bad bot networks**, or identifying bad bots that pretend to be human.

Similarly, one-off projects to solve computational propaganda are developing sporadically around the world.

Indiana University researchers are among institutions releasing beta versions of projects like OSoMe: Social Media Observatory. IU's OSoMe includes tools for **"visualizing the spread of claims and fact checking"** and "detecting and blocking Twitter bots in a newsfeed".

Ukraine's Kyiv-Mohyla Journalism School, along with the KMA Digital Future of Journalism, launched the Stopfake.org **fact-checking site**. The site is an information hub where users can examine and analyze all aspects of Kremlin-born online propaganda.

Government regulation & national security

Regulations to stop foreign political-propaganda online are also nascent.

The **Honest Ads Act** (H.R. 4077) is a bill introduced last year in the US Senate. The bill seeks to regulate online campaign advertisements hosted on platforms such as Facebook and Google.

In the EU, the **General Data Protection Regulation (GDPR)**, which goes into effect in May 2018, gives EU citizens the ability to control the use of their personally identifiable information (PII), which could help keep that data out of the hands of propagandists.

Ideas also include a call for the creation of a **new intelligence discipline** called "public intelligence." In theory, a new unit could "inform the U.S. public of hostile foreign activity intended to change beliefs or knowledge to the benefit of a foreign state."

Big tech is increasingly seen as not just an economic asset, but as a national security issue.

This changes the dynamics of how tech companies are viewed politically and puts pressure on these companies to act. Some tech giants are taking steps to combat the spread of so-called "fake news." For example, Google plans to spend $300M over the next three years to support authoritative journalism, while Apple recently acquired the digital magazine distributor Texture as an entry to the journalism world.

Final thoughts

The future of combating information warfare is uncertain but hopeful. The powerful cohort of DARPA, corporations, startups, non-profits, and universities are all making progress in the long-term fight against information warfare. Still, regulators and corporations alike have their work cut out for them.

Certain groups bare more of the brunt of responsibility for protecting individuals in the near-term. While corporations like Facebook are not directly creating propaganda, social networks do own and administer the pathways through which much of the world's digital deceptions flow. These platforms are under pressure to re-tool their sharing algorithms and business models to preserve users' privacy and help control the spread of propaganda and deception.

As regulators catch up to new technology and tech companies work to weed out bad actors, the onus for digital protection ultimately lies on the user. For the foreseeable future, we we will continue to be responsible for evaluating the truthfulness of the information we consume.

This means being aware of the narratives we are served, and teaching ourselves to recognize narratives that confirm our bias vs those that challenge our beliefs. But it's unlikely that individuals en masse will become independent information warriors capable of evaluating competing narratives with equal scrutiny. Barring this, other institutions will need to lay the groundwork for a scalable defense against information warfare.